Journal of the Association for Information Systems (2020) **21**(1), 90-116
doi: 10.17705/1jais.00593

RESEARCH ARTICLE

ISSN 1536-9323

Developing Theory Through Integrating Human and Machine Pattern Recognition

Aron Lindberg[1]
[1]Stevens Institute of Technology, USA, aron.lindberg@stevens.edu

Abstract

New forms of digital trace data are becoming ubiquitous. Traditional methods of qualitative research that aim at developing theory, however, are often overwhelmed by the sheer volume of such data. To remedy this situation, qualitative researchers can engage not only with digital traces, but also with computational tools that are increasingly able to model digital trace data in ways that support the process of developing theory. To facilitate such research, this paper crafts a research design framework based on the philosophical tradition of pragmatism, which provides intellectual tools for dealing with multifaceted digital trace data, and offers an abductive analysis approach suitable for leveraging both human and machine pattern recognition. This framework provides opportunities for researchers to engage with digital traces and computational tools in a way that is sensitive to qualitative researchers' concerns about theory development. The paper concludes by showing how this framework puts human imaginative capacities at the center of the push for qualitative researchers to engage with computational tools and digital traces.

Keywords: Digital Trace Data, Theory Development, Computational Tools, Pragmatism, Abduction, Human Pattern Recognition, Machine Pattern Recognition.

Suprateek Sarker was the accepting senior editor. This research article was submitted on March 31, 2017, and underwent three revisions.

1 Introduction

Digital trace data (Hedman, Srinivasan, & Lindgren, 2013; Shmueli & Koppius, 2011) are becoming ubiquitous. Such data represent residues left behind by multiple types of behavior that are collected by digital platforms, making it possible to follow the things that human beings do as they engage in various social, technical, and hedonistic activities (Hedman et al., 2013; Howison, Wiggins, & Crowston, 2011; Venturini & Latour, 2010). Further, such data do not come readily defined as operationalizations of concepts, as, for example, survey data do. Indeed, such data are "increasingly heterogeneous and unstructured—text, images, video—often emanating from networks with complex relationships between their entities." (Vasant Dhar, 2013, p. 64). Digital traces can be manipulated, transformed, and interpreted in multiple ways and are therefore inherently mutable (Shmueli & Koppius, 2011). As such, digital trace data represent a rich fount of raw material for qualitative scholars (Berente, Seidel, & Safadi, 2019; Vaast & Walsham, 2011; Walsh, 2015).

Qualitative researchers have traditionally worked to inductively develop theories that are sensitive to both context and the ways that individuals and groups constitute their social worlds (Charmaz, 2014). The standard method for achieving this has been to manually code text-data, such as interview transcripts and documents, so as to identify codes, themes, and concepts that are "grounded" in the data (Strauss & Corbin, 1990). Such coding is normally done as part of a process of theoretical sampling (i.e., purposeful sampling based on interim results) and constant comparison of all data until theoretical saturation is reached. Such methods excel at developing theory that is deeply grounded in multiple aspects of rich datasets, theories that also speak to how individuals construct their social worlds within particular contexts.

When confronted with digital trace data, however, such methods tend to become overwhelmed by the sheer size of some datasets. Digital traces can easily comprise thousands if not millions of data points, making it quite difficult for qualitative researchers to approach such data using manual coding techniques. For example, following principles of theoretical sampling and aiming for saturation (Charmaz, 2014, p. 214) when studying an online community with thousands (if not millions) of members, is likely to exhaust the resources of even the most well-funded qualitative research team. Further, certain structures, such as relationships between activities within a process, or relationships between individuals within a community, may be difficult to discover unless the totality of available data is examined, using tools that enable the identification of latent patterns that may not be obvious to the human eye (Lindberg, Gaskin, Berente, Lyytinen, & Yoo, 2013).

This leaves qualitative researchers in a quandary. There are new rich sources of data available, but these sources of data tend to overwhelm the traditional techniques that qualitative researchers have traditionally used to develop theory. To overcome this situation, my proposal is for qualitative researchers to examine a trend attending the rise of digital trace data, i.e., the emergence of computational tools, meaning specific analysis technologies, such as social network analysis and sequence analysis. The combination of digital traces and computational tools enables the use of unobtrusive data (Webb & Weick, 1979) to discover large-scale patterns that may complement traditional qualitative research (Whelan et al., 2016). Computational tools facilitate the discovery of structures and patterns across large datasets, a task that can be difficult, if not impossible for human beings to perform (Berente et al., 2019). Such tools are not necessarily tightly associated with traditional, positivist, hypothesis-testing approaches to research (Lee, 1991), but are in many ways intrinsically inductive in their approach to data.

Indeed, if induction is taken at face value—as attempts to identify regularities in how human beings socially construct their worlds in relation to their context—then computational tools are increasingly capable of doing this as well. Sequence analysis, text analysis, and social network analysis can all offer insights into the practices, understandings, and structures of relationships that shape how social worlds emerge, evolve, and stabilize over time. These are all inductive methods that find patterns in data, enabling humans to use those patterns to say something about the social world and how it has been constructed. Hence, both humans and machines have capacities for pattern recognition (Holland et al., 1989). While human pattern recognition (i.e., the capacity of the human mind to discern regularities in data) has long been central to efforts to develop theory, the capacities of machines to identify patterns (i.e., algorithms that identify regularities in data) usable in theory development are becoming increasingly dynamic, contextual, and sophisticated. Therefore, such tools represent an opportunity for qualitative researchers to extend their arsenal as they seek to continuously generate novel and insightful theories.

These opportunities allow qualitative researchers to expand but not necessarily replace their current toolboxes (Grimmer & Stewart, 2013). Because both humans and machines excel at different forms of pattern recognition (i.e., the identification of regularities in data) this expansion has the capacity to considerably broaden the capacity of qualitative researchers to investigate a fuller range of phenomena and their constituent aspects. Humans, using their sensemaking capacities (Weick, 1979), can, and inevitably will, place a conceptual, socially constructed layer on top of observed events and phenomena (Berger & Luckmann, 1967). When human beings draw upon networks of associations in order to contextualize and make sense of an observed event or speech act, they are actively associating an observation with other observations and with mental entities that relate to the observation, such as values, principles, mental constructions of causality, implicitly held theories, etc. All of these activities constitute human pattern recognition.

Machines have their own way of conducting pattern recognition—they can compute correlations across data points and use various statistical techniques to identify complex patterns and structures. Such computations are bundled in various modeling techniques, such as social network analysis, sequence analysis, or text mining. Each technique optimizes for a specific type of data, and specific types of relationships across data points (Džeroski, Langley, & Todorovski, 2007). Because of ever-increasing computing power and storage capacity, machine pattern recognition often excels at estimating such models across vast datasets.

Digital trace data and computational tools therefore present important opportunities for qualitative researchers. Digital trace data form a new source of rich, dynamic data that cover both human expression in text and traces of various behaviors. Computational tools offer opportunities to analyze large-scale patterns that are often difficult to capture using traditional qualitative methods. The main thrust of this paper is an insistence that qualitative scholars cannot ignore these developments; rather, they must engage seriously with these types of data and tools to maintain their relevance in an emerging world of rich data and sophisticated computational tools. This paper serves as a guide for qualitative researchers, demonstrating how to integrate digital traces and computational tools into their arsenals.

Supporting qualitative researchers wishing to engage with digital traces and computational tools in order to develop theory necessitates an epistemological perspective that allows for engagement with the dynamic, heterogeneous characteristics of digital trace data, as well as the multitude of traditional modes of qualitative analysis, along with new, emerging computational modes of analysis. To tackle this task effectively, I turn to the pragmatist tradition, originating with the American philosophers James, Peirce, Dewey, Mead (Whitford & Zirpoli, 2014), and more recently Rescher, Rorty, and Putnam (Rescher, 2000), as well as some European strands of this tradition (Latour, 2006; Venturini & Latour, 2010). These thinkers provide a diverse set of ideas that are uniquely suited to the task of approaching digital trace data and computational tools with the intent of developing theory, which stems from pragmatism's focus on action as a source of meaning and understanding of the social world. Because digital traces are records of human action, the pragmatist perspective can use such traces as a window into human nature and the social world, thus enabling the development of theory.

2 The Pragmatist Perspective

Pragmatist philosophy is driven by a disillusionment with ontological preoccupations (James, 1907). Rather than focusing on establishing truth claims through correspondence with some "reality" (either existing objectively or being socially constructed), pragmatists put action at the center of their philosophies. This means that inquiry, for example, is seen not as abstract cogitation, but rather as action that serves to elicit certain effects, or favorable consequences, thus validating that the inquiry has hit upon something of value (Dewey, 1938b). Human beings interface with the world not through disembodied cognition but through contextually embedded performance of action. The way that humans discover problems, negotiate solutions, and progress is therefore through action. Consequently, pragmatists such as James (1907) tend to define "truth" as whatever helps an actor generate "good." In contemporary terms, one would say that pragmatists are interested in "utility," meaning that knowledge is important to the degree to which it helps actors take action in the world to achieve specific goals. Note that this does not equate to a laissez-faire "whatever works is true" type of epistemology. Rather, inquiry into the consequences of action is conducted within a community (Dewey, 1938) of scientists (Ormerod, 2006). This focus on the consequences of action is therefore related to "scientific instrumentalism"—the idea that prediction is at the core of the scientific enterprise (Popper, 1965).

Due to pragmatism sidestepping ontology and focusing on action and its consequences, the abundance of digital trace data currently available offers an opportunity for qualitative researchers to engage in constructive and dynamic ways with such data as well as with computational tools. A number of pragmatist principles facilitate this engagement—specifically, the rooting of habits in agency (constitution), the embedding of action in specific situations and environments (context), and the centrality of causality to inquiry (consequences). Each of these aspects are explained below (see Table 1 for a summary).

First, the pragmatist view of action largely rests on the idea of action becoming habitual (Baldwin, 1988; Gronow, 2012; Winter, 2013) and therefore entrained both at the individual and the social level. This means that human beings develop propensities to act in specific ways, thus laying the groundwork for routines (Cohen, 2007; Cohen et al., 1996; Cohen, 2012) and capabilities (Eisenhardt & Martin, 2000) to emerge.

Table 1. Principles of Pragmatism

Principle	Description	Example
Constitution	Habits emerge from and become constituted by idiosyncratic action, while still maintaining the creative engagement with reality entailed by agency	Performance of routines constitutes their ostensive aspect
Context	The efficacy of action is contextually embedded and can only be made sense of within this context	The same course of action will produce different effects in different contexts
Consequences	The outcomes of actions are indicators of useful knowledge, as they effectively illustrate the "meaning" or "utility" of an action	Our understanding of an object is intimately tied to how we intend to use it

Still, the pragmatist view of action is different from the dominant strains of practice theory that draw upon Giddens's (1984) structuration theory, which have mostly powered various qualitative studies (e.g., Orlikowski, 2000) and, more recently, the application of critical realism (Wynn & Williams, 2012; Zachariadis, Scott, & Barrett, 2013) focusing on identifying generative mechanisms that produce structural patterns.

Structuration theory has a tendency toward isomorphism between the agentic and structural levels (Sewell 1992), whereas the pragmatist outlook strongly emphasizes the agency and creativity of individual actors confronting a specific situation embedded in a particular context. Critical realism tends to focus on explaining the emergence of structure through theorizing unobservable "generative mechanisms" to the exclusion of other theoretical concerns. A pragmatist view of digital trace data allows for the crafting of accounts that show, empirically, *how* patterns emerge from the idiosyncrasies of agentic action (Venturini & Latour, 2010), thus avoiding excessively abstract, structural, and oversocialized accounts of human practices (Granovetter, 1985) devoid of individual agency.

Second, pragmatic approaches, especially those drawing upon Dewey (1938,) maintain that actions can only be understood in terms of their context and the associated meanings attached to various actions (Burks, 1946). Action is always executed by someone in a specific situation under certain conditions. For example, Carlile (2002) shows how knowledge is localized, invested in action, and evaluated based on practical consequences. This is an important insight, because as much as pragmatism emphasizes causation, the concern is less about establishing general laws or patterns than about illustrating contextually efficacious practices (Farjoun, Ansell, & Boin, 2015).

Third, pragmatists are fundamentally concerned with evaluating the meaning of actions in terms of their consequences. It is hard to observe internal emotional or cognitive states, but it is possible to clearly observe actions and the consequences that such actions engender. Thus, when trying to understand how actors think about and interpret their worlds, it is necessary to also look at their actions and the consequences of those actions. Traditionally, some qualitative researchers have been reluctant to embrace causality (e.g., Orlikowski, 2000), but the pragmatist approach posits that understanding causality is central to understanding *meaning*, since the meaning of an action (or utterance, i.e., a speech act) largely resides in its consequences. Rescher (2000, p. 9) provides the following example:

> *Take the concept of an "apple" for example. When we characterize something as an apple, we commit ourselves to treating it in certain ways—to handle it, store it, use it, discuss it, and so forth in the particular way appropriate to apples. And this is what it means to be an apple.*

In a nutshell, the meaning of a concept is intertwined with our usage, intended or actual, of the concept or its referents.

In summary, pragmatism highlights that human action, consequences, and structures are situated in contexts. These tendencies, in combination with an openness to diverse methods, enable the examination of human meaning and behavior from multiple angles, thus moving beyond traditional forms of qualitative research. The type of research that emerges from these epistemological implications and their potential is what Pollock and Williams (2008) call "third-wave" studies—research that examines both rich contexts and abstracted structures and therefore enables the study of phenomena as they emerge from the micro- to the macro-level (Venturini & Latour, 2010). To understand further how the pragmatist method can actually be leveraged in a practical research situation, I next explain the heart of pragmatist inquiry: abduction.

3 Abductive Inquiry as Discovery and Justification

Pragmatism provides an analytical method that allows for the integration of the various facets of the pragmatic worldview with the mutable (meaning that they can be manipulated, transformed, and interpreted in multiple ways) digital traces at hand and the multiple human and machine pattern recognition approaches available. This method is called abduction (Paavola, 2005) and refers to the act of generating reasonable inferences that, if true, make sense of the data at hand. From the pragmatist perspective, abduction is conceived of as being broader than simple deductive, syllogistic inference, or induction from specifics to generalities. In short, abduction is a strategy (Sami, 2004), or a method of inquiry (Locke, Golden-Biddle, & Feldman, 2008), i.e., "the controlled and directed transformation of an indeterminate situation into one that is so determinate in its constituent distinctions and relations so as to convert the elements of an original situation into a unified whole" (Dewey, 1938, p. 108). As such, abduction is a method of inquiry geared towards dealing with "felt difficulties," which are described by Dewey as "cases of striking novelty or unusual perplexity, the difficulty, however, is likely to present itself at first as a shock, as emotional disturbance, as a more or less vague feeling of the unexpected, of something queer, strange, funny, or disconcerting" (1910, p. 50). Often, such situations occur when the confrontation between a theory and new empirical findings provokes a "breakdown," indicating a mismatch between theory and findings (Agar, 1985, p. 20.) To respond to this, new theory can be developed based on an iterative sensemaking process (Grolemund & Wickham, 2014.)

Traditionally, researchers think of the research task as roughly divided into two realms, which Swedberg (2012) calls the "context of discovery," i.e., the scientific practices that researchers make use of to generate insights, and the "context of justification," i.e., the validation of insights according to scientific principles such as falsification and adherence to the scientific method. The abductive approach, however, disrupts this neat division of scientific activities, since abduction is an activity that spans the boundary of both the context of discovery and the context of justification. In the context of pragmatist philosophy, abduction is often described as an iterative process whereby an analyst iterates across discovery and justification using empirical data while also making comparisons to extant and emergent theory.[1]

Using digital trace data, abduction starts with the discovery of patterns rather than the a priori formulation of hypotheses, because "patterns often emerge before the reasons for them are apparent" (Dhar & Chou, 2001, p. 907). Starting with either human or machine pattern recognition, inductive generalizations are generated. This process (captured in Figure 1 below) typically starts with machine pattern recognition if the priority is to analyze large-scale patterns and structures, and starts with human pattern recognition if the analyst wishes to build a foundation by zooming in on situated human dynamics such as agency and individual, lived experiences.

For example, Zachariadis et al. (2013) start with computational analysis of the overall relationship between IT implementation and banking performance, and then use qualitative inquiry to investigate the mechanisms that constitute this relationship. Hence, their study first identified a structural relationship and then inquired into the underlying mechanisms generating this relationship. In contrast, Miranda, Kim, and Summers (2015) started with the qualitative coding of how individuals express various aspects of "organizing visions." They then used this qualitative coding as input for a relational class analysis, a computational technique for eliciting relationships across constructs identified throughout a text corpus. Hence, in their study, individual, contextually embedded expressions of personal visions were elicited first, after which structural patterns were discovered. Thus, depending on whether an analyst wants to emphasize agency or structure, either human or machine pattern recognition may be more heavily emphasized throughout an analysis (Brown et al., 2016, p. 444).

Inductive generalizations could be computationally derived patterns, such as descriptive statistics and correlations, or other regularities, such as textual themes or categorizations of action. Such inductive generalizations then lead to the discovery of a "working hypothesis" of what patterns would be justified using, for example, human pattern recognition, that either explains or corroborates the inductive observations made using machine pattern recognition, or vice versa. These working hypotheses are "reasonable inferences" tempered by theoretical experience and intimacy with the data under scrutiny. Such reasonable inferences are drawn from the analysis of data but are not seen to be inductively or deductively true. Rather, they should be assessed on the basis of whether a conclusion drawn from the evidence is "reasonable," meaning that it *probably* follows from the data analysis. In this sense, such inferences often constitute a generalization from one empirical statement to another (Lee & Baskerville, 2003). The abductively generated working hypothesis is what explains the capacity of science to make new discoveries (Dougherty, 2016), rather than simply deducing testable propositions based on extant theory (Simpson, 2009). Such "imaginative leaps" are rooted in human instincts (Ayim, 1974) and openness to experience (Chiasson, 2007). A working hypothesis is thus a vehicle for sustaining the momentum of the inquiry process, about which Dewey (1938, pp. 144-145) states the following:

> *a hypothesis does not have to be true in order to be highly serviceable in the conduct of inquiry. Examination of the historical progress of science will show that the same thing holds good of "facts": of what has been taken in the past as evidential. They were serviceable, not because they were true or false, but because, when they were taken to be provisional working means of advancing investigation, they led to discovery of other facts which proved more relevant and more weighty.*

[1] Here the term "extant" theory is used to denote those theories that already exist within the literature, which are used as a theoretical background. The "emergent" theory is the result of the theorizing work done by the researcher as part of the abductive analysis process. Emergent theory serves as a tentatively and iteratively evolving explanation of the empirical findings elicited by abductive inquiry.

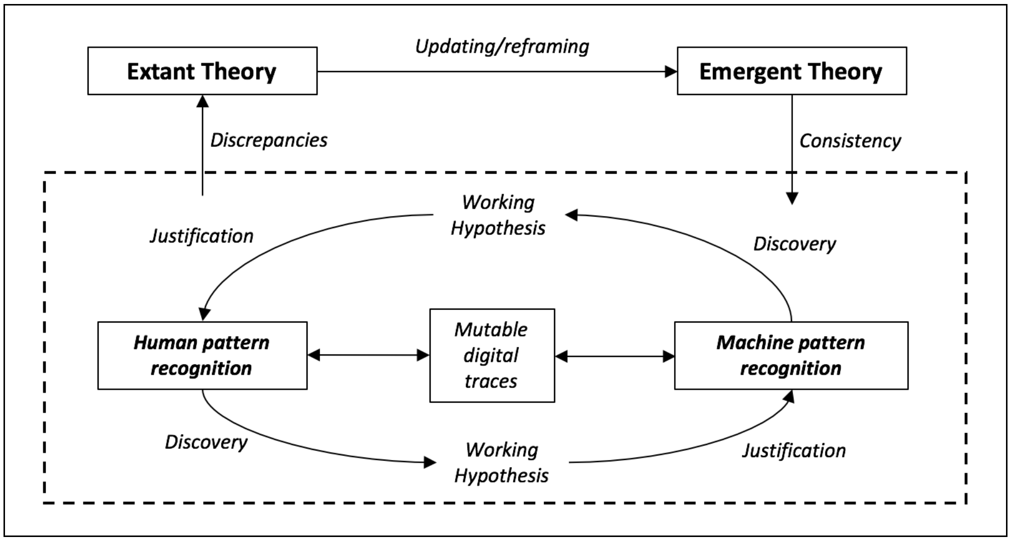

Figure 1. Abductive Inquiry as Discovery and Justification

Abduction thus consists of an iterative alternation between discovery and justification (see Figure 1). As reasonable working hypotheses are identified, they are also tested against other forms of data. For example, if a theme regarding, say, interpersonal conflict has been identified using human pattern recognition, a working hypothesis can be generated, specifying that we expect to see highly volatile activity patterns in conjunction with this theme. If such reasonable working hypotheses can be supported using machine pattern recognition, then there are grounds to accept the validity of both the theme and the pattern of behavior that have been identified. Effectively this amounts to a process of constant comparison between different forms of data (Charmaz, 2014), a process that also provokes the constant theoretical resampling of data in order to either discover or justify working hypotheses based on prior findings within the same analysis process.

Note that regardless of which type of analysis method is being used, abduction always draws upon the imaginative capacities of the human mind. This means that when analysis is conducted, whatever the mode of the analysis, neither the data nor the methods can speak for themselves. Rather, the abductive mode of inference requires an imaginative faculty (Paavola, 2005; Sami, 2004; Weick, 1989) that continuously creates inferences based on data and analysis. These inferences comprise not only deductive inferences but also "imaginative" inferences that imagine theoretical possibilities, as well as possible adjacent correlations, constitutive relationships, and causal processes. Hence, even when using machine pattern recognition for discovery, the active participation of the human, imaginative mind in generating theoretical propositions is necessary. This capacity helps humans forge explanations out of multiple, heterogeneous analyses in relation to extant and emergent theory with the intention of illuminating the phenomenon at hand. This is clearly a process that cannot be handled by machine pattern recognition. The imaginative faculties of humans are, therefore, essential for the act of (1) generating working hypotheses, and (2) comparing sets of justified working hypotheses with extant theory to develop emergent theory. Hence, the uniqueness of human beings in the overall process of theory development cannot be denied. Regarding raw pattern recognition across a delimited set of data points, however, human capacities can be augmented by those of machines (Grimmer & Stewart, 2013).

Alongside the iterations of human and machine pattern recognition, researchers compare empirical patterns to extant theory. The need for new, emergent theory grows to the degree to which findings identified by humans and machines are consistent with each other but discrepant with extant theory. Emergent theory is a way to craft an account or tell a "story" that makes sense of the data at hand (Grolemund & Wickham, 2014), i.e., to shape a new theoretical account that is consistent with the findings that have been identified and triangulated, whether a new theory represents an incremental improvement upon extant theory (i.e., updating) or a "rupturing" break (i.e., reframing) with extant theory (Walsh, 2015). The radical reframing or rupturing of theory is accomplished by problematizing extant theory (Alvesson & Sandberg, 2011) in light of empirical findings and formulating an emergent theory that can resolve problematic aspects of extant theory. That is, the theory that is being formulated must enable insights that are both "interesting" (Davis, 1971) and "important" (Rai, 2017). Hence, the emerging theory forms the "foreground," while the extant theory forms the "background." Through contrasting foreground and background, theoretical tension is achieved, meaning that the foreground appears in stark relief against the background, thus communicating its theoretical value—namely, new understanding and insight. This view of theoretical contribution is expressed by Weick (1995, p. 294) in the following way: "in a full defense [of an idea], the author shows how some display looks different before and after it is viewed using the innovation [i.e., the emergent theory] that is proposed."

4 Guidelines

In this section, I explain the guidelines for how to conduct research using the framework presented in the previous section. These guidelines are geared towards the qualitative researcher who wishes to introduce machine pattern recognition techniques into his or her work with the goal of developing theory. The guidelines are structured according to the two major aspects of the previously developed framework: mutable digital traces and abduction (discovery and justification). The guidelines are summarized in Table 2 below.

4.1 Guidelines: Mutable Digital Traces

The new "oil" of the digital era is data, and digital traces imply a way of conceiving of such data as tracks or remnants left by human action that remain after humans interact with platforms and other digital systems. Such traces, therefore, allow for a pragmatist perspective that emphasizes that human action is central to understanding how humans work, think, and feel. Below, I examine how qualitative researchers can approach digital traces and computational tools. Overall, I provide three general guidelines: sample data continuously to resolve emerging puzzles, maximize richness of digital trace data, and craft crisp constructs that move beyond the emic meaning of measurement.

4.1.1 Guideline #1: Sample Data Continuously to Resolve Emerging Puzzles

Sampling of data is not a once-and-done process, but rather a continuous search for data (Behfar & Okhuysen, 2018) that may be helpful in discovering and justifying working hypotheses. Such sampling may occur as a reaction to a posed (i.e., discovered) working hypothesis, thus prompting the collection of particular data necessary to justify the working hypothesis. Similarly, once a working hypothesis has been justified, it may suggest the existence of datasets that may potentially lead to the discovery of new working hypotheses, thus suggesting a "speculative search" (Behfar & Okhuysen, 2018) for new data, akin to what grounded theorists call "theoretical sampling" (Charmaz, 2014).

The continuous sampling of data to resolve emerging puzzles encourages researchers to move across all three pragmatist principles. For example, once an understanding of the constitution of a phenomenon, e.g., an organizational routine, has been developed, one might ask in what context this routine emerges. Once this process of emergence has been understood, one may then ask what the consequences of the routine are. Such movements across the different principles of pragmatism enable the resolution of emergent puzzles in a cyclical and iterative way.

As a further example, consider Zachariadis et al. (2013) who computationally identified a relationship between the implementation of a specific IT system and performance among banks in London, which then prompted further, qualitative data collection to elicit a mechanism through which the observed relationship was constructed. Here, the structure of the identified relationship provided guidance with regard to the type of data that should be collected. The researchers did not merely attempt to "triangulate" across various data sources and analytical methods, but rather used relationships identified in one form of data (digital traces) using a particular analytical technique (regression) as inputs into determining the next step of the analysis (qualitative analysis of interview data).

4.1.2 Guideline #2: Maximize Richness of Digital Trace Data

In order to prepare for analyses that cover all three pragmatist principles, analyzed using both human and machine pattern recognition methods, datasets need to be collected that are varied and rich. Such data collection therefore constitutes a process of "data expansion" (Behfar & Okhuysen, 2018, p. 332) through which multiple views of a phenomenon are captured. Data that can be construed in multiple different ways are therefore important to the abductive analysis process. Collecting such data may also support the quality of inferences made later on, due to "sample integration" (Brown et al., 2016). Importantly, this enables analyses drawing on the principle of constitution, which often requires multiple types of data to both zoom in and out of (Gaskin et al., 2014). In particular, it is important that such data are rich along the following vectors: text, categories, discrete units, time stamps, and actors. Below I discuss each of these vectors in turn.

First, it is vital that the data have a textual component. The textual component allows for insight into the thought-worlds and lived experiences of the people who have left particular digital traces behind them as they interacted with a particular digital platform or system. Such text can be used by qualitative analysts to construct narratives, but it can also be used by machine pattern recognition through, for example, the application of text mining methods. Often, the qualitative interpretation of such textual data can serve to generate a deep contextual understanding of a particular phenomenon, thus providing a "binding glue" that ties together all the other components of an analysis.

Second, the data should also include categorical variables. This is often based on taxonomies specified by the platform itself, i.e., categorizations of various actions, such as "commenting," "posting," "adding," or "deleting" content. Such categories can be helpful for tracing processes of various kinds using longitudinal methods such as sequence analysis.

Third, it is helpful if such data are organized into discrete units. This helps to structure analyses of processes, sequences, or narratives by, for example, enabling the construction of qualitative narratives through identifying different "speech acts" (Searle, 1969). Forum data, for instance, are suitable for this purpose, because they make it possible to break the text down into units that have different posters and time stamps attached to them, which enables the construction of social networks, sequences, or textually based narratives, thus facilitating better analysis of the mutual constitution of agency and structure.

Fourth, through clearly showing the order in which things are structured, time-stamped data enables the construction of narratives through human pattern recognition, as well as the statistical estimation of processes and their attendant characteristics, through machine pattern recognition. Time stamps help the analyst adopt a rigorous frame of mind with regard to both the ordering (what precedes what) and the pacing (how much time passes between events) of events (Howison & Crowston, 2014; Lindberg et al., 2016). Based on such data, narratives can be constructed that follow tightly along categorical variables, thus laying the groundwork for tightly integrated analysis and theory development. Such data are crucial for addressing the pragmatist principle of consequences.

Fifth, it is helpful if data are stamped by the actor executing a particular action or posting a particular comment. This helps to identify relational dynamics while reading text (i.e., human pattern recognition) but also enables machine pattern recognition to build social networks that can be analyzed computationally. For example, if two actors perform actions or post comments within the same workflow, we can make the assumption that they are somehow connected to each other, i.e., their relationship can be characterized as "working together on issue X" (Howison et al., 2011).

4.1.3 Guideline #3: Craft Crisp Constructs That Move Beyond the Emic Meaning of Measurement

The crafting of constructs using human pattern recognition must be done with an eye toward computational crafting of adjacent concepts, corollaries, or operationalizations (see, e.g., Lindberg et al., 2016). This increases the pressure on the qualitative analyst to create constructs that have firm boundaries and describe aspects of activity processes, relationships, or textual themes in a manner that allows those qualitative constructs to be connected to computationally crafted constructs (Goertz, 2006). Qualitative identification of constructs must therefore occur in dialogue with the computational identification of constructs, thus preparing them for alignment with each other. This means that qualitative constructs must be clearly defined and have clear boundaries in terms of which processes, relationships, or textual themes they relate to, under what conditions, and in what subsets of the data. Constructs that are discrete, binary, or have clearly identified continua tend to be more helpful, as compared to constructs that are ambiguous, cover wide domains, or are impressionistic in nature. Such forms of conceptualization have been captured by "pretheoretic lexica" that help researchers frame unwieldy and ambiguous data according to clearly definable categories, rules, or continua (Berente et al., 2019) and often form a necessary first step on the path toward developing theory.

Creating constructs based on machine pattern recognition must be done in a way that allows them to be translated into usage by humans applying their innate pattern recognition capabilities. This means, for example, that the quantitative urge of rendering everything into continuous variables must be resisted. As such, constructs may refer to processes, or constellations of relationships, values, or interpretations of the social world. These may not be fully "deduced" from computational analyses, but they may very well be suggested. Crafting such constructs requires movement beyond the familiarity of the types of variables and measurements that tend to be found in digital trace data. While these measures tend to lack researcher bias, simply because they are defined by platform designers and not researchers, this does not mean that they are fully realized stand-ins for the concepts in which researchers are actually interested.

Rather, researchers need to actively work to raise the conceptual height of thinking regarding what such traces may actually indicate, i.e., identify the concepts that actually interest them and locate the degree to which various combinations of available digital traces may serve as indicators of these concepts, rather than simply accepting the digital traces at face value (Webb & Weick, 1979). For example, a set of markers may capture specific forms of contributions on a digital platform (e.g., posting comments, editing content), but one can view these contributions as indicators of higher-level processes, such as coordination or socialization processes. Raising the conceptual height of conceptual and theoretical development activities may therefore serve to create richer working hypotheses that may interrelate the principles of constitution, context, and consequences, instead of simply engaging in "dustbowl" empirical pattern matching.

Table 2. Guidelines

Aspect	Guidelines	Description	Relationship to pragmatist principles	Examples
Mutable digital traces	#1: Sample data continuously to resolve emerging puzzles	Data need to be pertinent to the research problem, but are also a reaction to prior puzzles uncovered during the analysis process	Encourages movement across principles	Characterizing a relationship between two variables in quantitative terms may suggest the gathering of qualitative data to explain the nature of the relationship
	#2: Maximize richness of digital trace data	Data need to contain both text, categories, discrete units, time stamps, and actors	Prepares for analysis of all three principles	Workflow data can be collected in ways that preserve both text (e.g., comments), as well as categorical variables such as time-stamped activity types and associated actors, all organized in discrete units
	#3: Craft crisp constructs that move beyond the emic meaning of measurement	Constructs based on indicators found in digital traces need to be fashioned in such ways that sensible corollaries elicited by either human or machine pattern recognition are enabled	Encourages interrelating principles	Qualitative categorizations of relationships in terms of intensity and characteristics can be complemented by social network-based metrics
Discovery and justification	#4: Search for explanations to surprises	Inconsistencies between extant theory and patterns elicited from data are the key driver of abductive inquiry	Principles facilitate comparisons between extant and emergent theory	Nonlinear effects may be detected where prior theorizing has only described linear effects, thus guiding the search for explanations
	#5: Zoom in and out	Zoom in on contextually embedded action performed by individual agents, while also zooming out to examine the effects of structures and contexts	Draws upon the principle of constitution	Simultaneous identification of structural patterns (using, e.g., graph-, sequence-, and text-analyses) and individual behavioral rules using qualitative analyses
	#6: Trace nomological networks	Trace nomological networks through correlations and modeling techniques	Draws upon the principle of context	Nomological networks can be extracted through looking at quantitative correlations, various quantitative modeling approaches, or through theoretical sorting, diagramming and integration of qualitative memos
	#7: Model causality	Model cause and effect relationships to expose processes	Draws upon the principle of consequences	Regression modeling and longitudinal panel methods may be used to establish causation through correlation, temporal precedence, and controls. Further, processes and narratives can be traced using qualitative coding to identify causal relationships. Finally, visualizations may be used to identify interactions across multiple variables within a process.

4.2 Guidelines: Discovery and Justification

The abductive mode of analysis is performed through discovery and justification, which directly apply the principles of pragmatism. While both humans and machines have the capacity to recognize a wide variety of patterns, this is not an automatic process; it is a process that requires focus, as well as an understanding of the foreground and background of various aspects of the reality to be examined. To that end, I provide guidance to qualitative scholars wishing to integrate machine pattern recognition into their research by providing four additional guidelines: search for explanations to surprises, zoom in and out, trace nomological networks, and model causality.

4.2.1 Guideline #4: Search for Explanations to Surprises

Identifying potentially fruitful working hypotheses is most often a matter of looking for what is out of place, inconsistent, surprising, and therefore interesting to the human mind (Davis, 1971). That is, inquiry is sparked by an "indeterminate" situation (Dewey, 1938). The abductive process then consists of an iterative "moving back and forth between the observed facts and the conditional idea…till a coherent experience of an object is substituted for the experience of conflicting details" (Dewey, 1910, p. 83). To resolve such situations, reasonable working hypotheses must be "suggested:"

> *Suggestion is the very heart of inference; it involves going from what is present to something absent. Hence, it is more or less speculative, adventurous. Since inference goes beyond what is actually present, it involves a leap, a jump, the propriety of which cannot be absolutely warranted in advance, no matter what precautions be taken.* (Dewey, 1910, p. 75)

Therefore, the foundation of pragmatist inquiry is to search for explanations to surprising observations. This practice permeates the entire research project, from design to theorizing, thus ensuring that all the elements of the research contribute to resolving a particular "indeterminate" situation. The search for such explanations to surprises is, in itself, guided by the three pragmatist principles—inquiring into the constitution of a phenomenon, the context in which it occurs, and its consequences—which provide theoretical "hints" regarding where fruitful inquiry is likely to occur. Drawing upon these principles therefore helps identify extant theories, to which emergent theory can then be contrasted.

4.2.2 Guideline #5: Zoom In and Out

Drawing on the principle of constitution, both the structures in which action may result, as well as the structures in which action is embedded, are relevant. Structures consist of accumulated patterns of microlevel behaviors that provide routines, habits, and practices to follow. Analyzing such multilevel data essentially amounts to what Gaskin et al. (2014) call "zooming in and out," i.e., the act of moving between both abstracted structures and contextually embedded analyses in a single study. Through utilizing human pattern recognition capacities, often in relation to text, an analyst can examine the contextually situated dynamics of individual agents. Then, using computational tools, which excel at finding structure in large volumes of data, the analyst can "zoom out" to see the larger picture in which agentic action is embedded.

For example, in literature studies, the standard for critics has long been "close reading" by humans who interpret the meaning of a text and put it in relation to other phenomena or viewpoints (Hirsch, 1967). However, with the rise of computational tools, some literary scholars have turned to text mining as a means for "distant reading," i.e., identifying large-scale patterns across large volumes of books or other writings (Moretti, 2013). Close reading is, thus, a more traditional qualitative technique, while distant reading is a computationally enabled technique. Both of these techniques can work together so that both deeply contextually embedded action and large-scale structural patterns become visible.

Therefore, utilizing the increasing dynamism of computational tools, qualitative analysts can complement their traditionally intensive, idiographic analyses of specific instances with computational tools capable of revealing the larger structural patterns in which such instances are embedded. For example, graph-, sequence-, and text-based analyses can be used to identify structural patterns across large swaths of data. Graph analyses are most commonly used to identify social structures and positions (Wasserman & Faust, 1994), but can also be used to identify relationships across artifacts or across humans and artifacts (Contractor & Monge, 2011). Similarly, sequence analysis can be used to analyze the structures of routines (Pentland, 2003) or sociomaterial processes (Gaskin et al., 2014). Finally, text-based analyses can be used to identify latent patterns in large bodies of text (Grimmer & Stewart, 2013).

The relationship between structure and agency can then be analyzed by, for instance, agent-based models that allow for the examination of the process through which specific agentic attitudes and behavior in the aggregate lead to the emergence of structural patterns (Bonabeau, 2002; Holland, 1992). Such models, which capture individual behavior with high degrees of precision, can effectively draw on ethnographic research methods to identify the behavioral rules of individual agents (Tubaro

& Casilli, 2010). Qualitative methods thus work in concert with machine pattern recognition methods, supplying the latter with inputs for structural modeling and simulations that may help clarify interactions between the micro- and macro-levels of a phenomenon.

4.2.3 Guideline #6: Trace Nomological Networks

The principle of context helps analysts probe the context in which an observed action occurs and helps them investigate various actions that are likely to occur in a particular context. Context constitutes adjacent information that contributes to making sense of an event, utterance, or observation. Hence, pragmatist philosophy provides tangible guidance regarding where working hypotheses can be discovered and justified through essentially looking for the context in which an action would reasonably occur, as well as looking for the context that may explain an action. This is illustrated by the familiar principle of searching along the nomological network (Cronbach & Meehl, 1955) of a concept, which implies the identification of correlations with other variables that are to be expected.

Berente et al. (2019, p. 53) discusses a similar practice, which they call "synchronic analysis," i.e., "identification of concepts and associations in any given moment in time," and note that such analysis can be performed by either qualitative or computational means. Using human pattern recognition, this may be accomplished through interrelating various codes, themes, and concepts that emerge from textual analysis. For example, in the grounded paradigm, such tracing of connections across concepts is accomplished through theoretical sorting, diagramming, and integrating of memos that the analyst has written during the iterative process of data collection and coding (Charmaz, 2014, pp. 216-224). Further, tracing relationships within a nomological network can also be conducted using machine pattern recognition through, for example, statistical modeling techniques such as factor analysis, regression, and structural equation modeling (Hair et al., 1998).

4.2.4 Guideline #7: Model Causality

Pragmatism directs researchers to focus on action—things that people are doing and the consequences thereof. The establishment of causality is, according to some perspectives, crucial to the establishment of theory. Berente et al. (2019, p. 53) argue that such forms of "diachronic analysis," i.e., the "identification of time-dependent relationships between concepts, for instance, in terms of cause-effect relationships" are the basis for moving beyond simply eliciting constructs to connecting them in a way that helps make sense of causal relationships.

This can be accomplished using both human and machine pattern recognition. From the perspective of the former, action is often captured by practices (Savigny, Schatzki, & von Savigny, 2001). The pragmatist perspective also encourages the examination of the outcomes of such practices in terms of performance, consequences, generated artifacts, or other desirable outcomes that might be of interest. To do this, the qualitative analyst may, for example, utilize process-tracing (Langley, 1999) to map out the particular ways in which outcomes are arrived at. Process tracing can be accomplished in multiple ways, including explicating events as rich "case stories" (Wynn & Williams, 2012), through visual means[2] (Tufte, 2001), through eliciting narratives, or through temporal bracketing (Langley, 1999, p. 696).

From a machine pattern recognition perspective, various approaches such as longitudinal regression and econometric panel methods (Bates et al., 2015) can be utilized to establish causation in ways that clearly respect conditions of correlation, temporal precedence, and controlling for spurious causes. Other, more sophisticated techniques include variable-length Markov chains (Bühlmann et al., 1999), temporal qualitative comparative analyses (Ragin & Strand, 2008), or process mining (van der Aalst et al., 2011). Each of these can be used to examine how specific sequential combinations of activities or practices of various kinds lead to particular outcomes.

5 Evaluating Abductive Inquiry as Process and Product

The strength of using digital traces and computational tools as a complement to human pattern recognition capacities is, from the perspective of the qualitative researcher, that a tight integration across multiple forms of data and analytical devices can be achieved. Through multiple forms of data, analyzed in multiple ways, a rich and tightly integrated account of empirical events exhibiting high degrees of validity can be achieved. Since pragmatist abductive inquiry places strong emphasis on the iterative relationship between the analytical process and its theoretical product, it is necessary to evaluate both the *process* and the *product* of such research (Behfar & Okhuysen, 2018, pp. 334-336). Below, I discuss each of the ways of evaluating research conducted in the form suggested in this paper (summarized in Table 3).

[2] For example, Tufte (2001, p. 41) provides a reproduction of Charles Joseph Minard's visualization of Napoleon's catastrophic Russian campaign of 1812. The visualization combines the temporal and geographic movements of Napoleon's army, along with changes in temperature. The core of the visualization is an increasingly thin line, indicating the decreasing size of Napoleon's army during both advancement and retreat. Hence, we can start to discern the suffering of the soldiers caused by battles, low temperatures, and extended travel on foot.

Table 3. Evaluating Abductive Inquiry as Process and Product

Aspect	Definition	Example
Process		
Problem	The fulcrum around which the data collection and analysis revolves	The problem guides which data should be collected, and what analysis may help to gain leverage over the problem
Data	Means for bridging problem and analysis and achieving sample integration	Collecting data that allows for all analyses pertinent to the problem to be performed
Analysis	Means for establishing coverage and connection	Analyses that cover the entire phenomenon under scrutiny, performed in such a way that the different analyzes can be connected to each other
Product		
Constitution and context	Situating practices within particular contexts	Routine dynamics may change depending on context
Context and consequences	Situating causal mechanisms within particular contexts	The same mechanism may have different consequences in offline/online contexts
Consequences and constitution	Identifying the causal consequences of social structures	Wiki governance structures may have specific causal consequences

5.1.1 Evaluating the Process

The evaluation of the process focuses on how tightly the research problem, data, and analytical techniques are integrated. As described above, abductive research flows from a surprising, baffling problem that demands resolution in an iterative manner. Hence, the more tightly that the problem, the data, the analysis, and the resulting theory are integrated, the more confident one can be that the work process itself has been rigorous. The integration, and therefore also the absence of "awkward fits," strengthens validity since different aspects of the analysis serve as checks and balances in relation to each other (Ketoviki & Mantere, 2010).

The problem, i.e., the central conundrum that research seeks to solve, forms a fulcrum around which both data and analysis revolve. As such, appropriate forms of both data and analysis need to be chosen in order to gain leverage over the problem. Within the framework presented in this paper, this involves working with different forms of data and analysis, all the while trying to make sure that the different forms of data and analysis are well-integrated with each other. I discuss this in terms of (1) how the data need to bridge the problem and the analyses, (2) how sample integration (Brown et al., 2016) can be achieved with regard to the data, and (3) how both "coverage" and "connection"[3] can be established with regard to multiple different analyses.

The guidelines in terms of the mutable traces discussed above serve to maximize both the richness and interpretative flexibility of the data collected, thus providing the analyst with more degrees of freedom in his or her work. Naturally, however, every dataset is not equally suited for solving every problem, so the data must be chosen in a way that captures pertinent aspects of the phenomenon in which the problem is grounded. This may involve a focus, for example, on a particular context, set of practices, or specific dynamics that a researcher seeks to examine in relation to a specific problem.

The data establish a bridge between the problem and the types of analysis that the researcher envisions will help clarify the problem. The chosen data, therefore, need to be both pertinent to the problem and capable of enabling analyses that may help resolve the problem, thus establishing a degree of integration across problem, data, and analysis.

[3] "Coverage" and "connection" represent a summarization of the categories in Table 1 from Venkatesh, Brown, and Bala (2013, p. 26), where complementarity, completeness, and compensation are sorted under coverage, while developmental, expansion, and corroboration/confirmation are sorted under connection.

The data must be appropriate for sample integration; thus, the researcher must ensure that the samples cover the same population (Brown et al., 2016)—i.e., it is critical that the data being used cover the same individuals performing the same activities. If this is not the case, the validity of the data in relationship to the problem being investigated may be lacking (Onwuegbuzie & Johnson, 2006). While the data to which human and machine pattern recognition are applied do not need to be identical (e.g., through applying both topic modeling and manual grounded theory to the same texts), the data must pertain to the same people, events, processes, or structures. If this is not the case, then the researcher is actually conducting multiple, separate studies dealing with the same general phenomena across different samples. Athough this may be valuable in some situations (see Berger and Pope [2011] for an excellent example of this approach), it does not allow for the multitude of interconnections across data and analyses that the pragmatist approach proposed in this paper affords. Throughout the analysis of the data, human and machine pattern recognition work together. This helps achieve coverage of the various aspects of the phenomena being investigated and establishes connections across these different aspects.

Seeking to provide "coverage" of the aspects of a phenomenon that are relevant to the resolution of a problem involves utilizing the different types of data that have been collected and representing them in multiple ways to enable multiple forms of analysis. Machine pattern recognition can often be used to elicit "thin" analyses stretching across large populations, while human pattern recognition often allows for conducting "thick" analyses over more limited ranges within a population (Geertz, 1973). For example, Lindberg et al. (2016) examined the "structure" of emergent routines using computational analyses, and then utilized qualitative content analysis to investigate the "content" of these routines, which allowed them to cover the full spectrum of the phenomenon pertaining to the research problem at hand. This suggests that humans need to interpret what correlations and models mean, i.e., they must seek to understand the implications of what people are actually doing, feeling, and thinking. Such interpretations serve two purposes: first, they help "flesh out" observations in terms of their significance within a particular social context, and second, they help generate implications, or in pragmatist terms: "reasonable working hypotheses," which may be tested using additional data.

Once such coverage has been achieved, the multiple aspects of a phenomenon need to be connected to enable a movement from conceptualization (i.e., a set of concepts) to theory (i.e., interrelated concepts). Establishing such connections implies a partnership between humans and machines. A clear example of this type of partnership between humans and machines is visualization. Machines can interrelate and plot data in various ways, while humans can easily and quickly identify visual patterns in such data plots. A further example is constituted by cluster analysis (Kaufman & Rousseeuw, 2005), which enables machines to identify groupings in large datasets that are hard for humans to discern, but which can then be interpreted by humans through inspection of the shared and differentiated attributes of these groupings.

When working hypotheses are generated on the basis of machine pattern recognition, these hypotheses tend to concern the meaning of patterns that have been identified, and human beings are able to situate those patterns within a larger context of principles, attitudes, expressions, as well as other observations that might not be immediately connected to a given model or visualization. Therefore, human pattern recognition can also serve to justify hypotheses, i.e., to explain how a certain pattern makes sense in relation to other patterns. Similarly, when human pattern recognition forms the basis of various working hypotheses, these hypotheses can be tested through machine pattern recognition, which is how statistical methods such as regression and other forms of modeling are traditionally used. Modeling, however, is not the only machine-based method of confirming hypotheses. Confirmatory evidence can also consist of visualizations, correlations, or other patterns that point at systematic relationships in the data, thus supporting observations that were made through human pattern recognition (see, e.g., Fayyad, Grinstein, & Wierse, 2002).

Therefore, working to establish both coverage and connection across multiple aspects of a phenomenon implies a process of cross-validation that draws on both human and machine pattern recognition. While digital trace data implies a rather positivist outlook on human behavior—we examine what we can observe (Abbott, 1992)—this is not to be taken as skepticism of interpretative studies or traditional qualitative inquiry into the dynamics exhibited by various social processes and forms of interaction. On the contrary, substantial relationships between agents' exhibited behavior and the experience of the agents enacting this behavior are to be expected. Such relationships are not necessarily direct correspondence relationships, but they are also unlikely to be random or disconnected relationships. Rather, we would expect to see complex bidirectional relationships, causal as well as constitutive, between the behavior exhibited by actors and their subjective experience of those behaviors.

5.1.2 Evaluating the Product

The product of abductive inquiry is a theory and can be evaluated based on how it integrates the various aspects of the pragmatist principles, i.e., to what degree constitution, context, and consequences are integrated into a seamless theory that addresses a particular research problem. This often takes the form of showing a process (consequences) that interacts with its environment (context), while at the same time also exhibiting the iterative dynamic between structure and agency (constitution). Such an account is a fully pragmatist account that is often enabled by a rich integration of analytical techniques and different forms of data. Evaluating how these principles are integrated with each other, as well as how they are used to address a particular problem, can clarify the product of the pragmatist, abductive process. Below, I examine how the three principles can interact with each other.

Practices and the structures that they form occur within particular contexts and pragmatism provokes an examination of this relationship. For example, in the theory of routines (Feldman & Pentland, 2003) routines are conceptualized as having two distinct aspects: performative and ostensive, where the former is a routine as it is actually performed and the latter is a formalized representation of this routine. These two dimensions are thus analogous to agency and structure. The interaction across them shapes the routine over time. Routines, however, do not exist in a vacuum; rather, they exist in a particular environment or context. This context also shapes the routine, which adapts to the environmental pressures that it is faced with (Aldrich & Ruef, 2006). Hence, the constitution of practices, habits, routines, and structures, needs to be examined in terms of how it interacts with the context in which these constitutive processes are embedded.

Similarly, context also interacts with the consequences of various processes. That is, any causal mechanism or process that yields particular outcomes occurs in a particular environment or context. For example, a common moderator used in studies based on digital trace data is the context of online communities (Faraj et al., 2016). Many theories have been developed based on studies of offline contexts situated within formal organizations, which has prompted scholars to ask whether such theories are moderated by context, e.g., if the same things occur in the same way in online communities. Hence, consequences of various causes need to be examined within the contexts in which they occur.

Finally, constitution interacts with consequences so that different structures have different causal consequences (Elder-Vass, 2011) on parts of sociotechnical systems that they do not directly constitute, but affect causally. For example, beyond exhibiting emergent dynamics, many online communities also have some instances of top-down governance schemes. Wikipedia, for example, has administrator role structures that can adjudicate conflicts of various kinds (Arazy et al., 2011). These are examples of social structures that have consequences on processes within a system, even if they do not directly constitute these processes. Integrating the examination of causality and constitution will contribute to a more well-integrated theoretical account.

Examining the interaction of these different principles facilitates judgments about how well-integrated a theoretical account is and how well it utilizes the different principles of the pragmatist worldview. These integrations must then be judged against extant theory so that a contrast appears. As research progresses, it becomes more and more difficult to identify new variables and processes which, by themselves, are novel. Therefore, examining interactions across multiple aspects of a theory is increasingly becoming a fruitful place for researchers to look when seeking opportunities to contribute (Locke & Golden-Biddle, 1997).

6 Three Empirical Examples

A number of researchers interested in information, technology, and organizing have already begun to analyze digital trace data in ways reminiscent of the approach proposed herein. These researchers often draw upon both human and machine pattern recognition and are concerned with developing theory that recognizes both constitution, context, and consequences. I review three such studies: Lindberg et al.'s (2016) study on emergent routines in open source software development, Tuertscher, Garud, and Kumaraswamy's (2014) study on coordination and knowledge interlacing at ATLAS/CERN as well as Vaast et al.'s (2017) study on patterns of microblogging in the wake of the Mexican Gulf Oil Spill. These studies are summarized in Table 4 below.

Published research papers represent well-organized "memoir-like" accounts of a research process and its eventual product, as opposed to messy "diary-like" accounts of all the twists and turns of the research process. The abductive approach is largely a way to guide a researcher through this messy process, which is often highly simplified in the published paper in order to aid reader comprehension. In my reading of these papers, therefore, I focus on evaluating the process (as described in the method section) and the product (as constituted by the proposed theory) along the lines suggested in the previous section. I evaluate the process based on its integration of problem, data, and analysis, and evaluate the product based on its integration of the overall research problem and the three pragmatist principles of constitution, context, and consequences.

Table 4. Examples of Research Integrating Human and Machine Pattern Recognition

Evaluation	Aspect	Lindberg et al. (2016)	Tuertscher et al. (2014)	Vaast et al. (2017)
Problem		How can OSS developers coordinate unresolved interdependencies, despite lacking traditional coordination mechanisms?	How can actors develop a novel, complex technological system despite the lack of hierarchy or a central coordinator?	How does social media afford connective action?
Process	**Data**	686 routine performances consisting of 3,707 activities, 432 text excerpts, and 17 interviews Sample integration is achieved through treating the same data (i.e., routine components) in multiple, different ways	328 meetings, 84 semi-structured interviews, 128,015 mailing list items, and 2,419 documents (meeting notes, etc.) Sample integration is achieved through collecting multiple forms of data from the same overall process	23,000 tweets on the Gulf of Mexico oil spill of which 1,882 tweets focused specifically on three specific "connective action episodes" Sample integration is achieved through treating the same data (i.e., tweets) in multiple, different ways
	Analysis	Utilizes multiple levels of coding, visualizations, and regressions to address the research problem from multiple angles Analyses cover both covariation patterns as well as qualitative categorization (i.e., content analysis) and establishes a connection across qualitative categories and other data using logit regression and ANCOVA Leans toward machine pattern recognition (sequence analysis); manual content analysis is used as input to machine pattern recognition	Utilizes qualitative coding to arrive at a process model, text mining to show patterns of justification and contestation, and graph analysis to capture interlaced knowledge Analyses cover knowledge distributions and forms of communication that connect these analyses through visualizations Leans toward human pattern recognition performed through grounded theory; machine pattern recognition is used in a confirmatory manner	Utilizes cluster analysis to identify roles and qualitative coding to identify episodes as well as confirm roles, graph-based motif analysis is then used to show particular interaction patterns across roles Analyses cover roles, connective action episodes, as well as social structure motifs, all of which are connected to each other through visualizations and cross-tabulation Leans toward machine pattern recognition (graph analysis and visualization of temporal patterns); human pattern recognition was used to frame the analysis
Product	**Constitution**	Routines emerge from local adaptations	Individual acts of contestation and justification constitute a larger coordination process	Global roles are enacted in specific situations
	Context	Different types of routine variation respond to different types of work-related interdependencies	Different intensities of justification across different contexts exhibiting differences in interlaced knowledge	Processes unfold across multiple, contextually different episodes
	Consequences	Activity variation in routine performances leads to merging	A process-based view of coordination	Different affordances enable different types of connective action

6.1 Lindberg et al. (2016)

In this study the authors explain how open source software (OSS) developers coordinate interdependencies among themselves and across the software code, despite lacking hierarchical organizing mechanisms. The authors mainly used digital trace data to develop a theory of coordination in the context of open source software development, using 686 routine performances consisting of 3,707 activities, as well as 432 text excerpts coded using open and axial coding. The digital trace data used are rich and could thus be treated in multiple ways (i.e., both qualitatively as well as in terms of categorical variables), therefore helping to establish sample integration.

To address the research problem, the data were analyzed in multiple ways. Exploratory data analysis (Tukey, 1977) was used to examine relationships across various interdependencies and different forms of routine variation. This initial analysis then served as the motivation for conducting a content analysis of workflows in order to elicit different components of routines: direct implementation and knowledge integration. These routine components were then connected to varying forms of routine variation. Finally, a regression showed that one form of routine variation, activity variation (essentially the diversity of activity types within a routine), helped predict whether code was successfully merged (i.e., accepted) into the codebase or not. The data and the analyses therefore emerged directly from the problem that the authors sought to address.

The analyses cover both various covariation patterns (established through visualizations, tests of mean differences, ANCOVA, and logit regressions) as well as qualitative theme identification (i.e., identification of routine components), thus helping to clarify both the structure and content of the work being conducted. These different analyses were connected with each other, most noticeably through the use of an ANCOVA. Table 6 on p. 761 shows how the qualitative themes are related to measures of activity- and order variation arrived at by the use of machine pattern recognition.

In terms of the balance between human and machine pattern recognition, this study leans mainly on machine pattern recognition (sequence analysis), while human pattern recognition (content analysis) was used as an input to regression modeling. Hence, this study mainly illustrates the use of computational tools and shows how they may be helpful for analyzing structural patterns in an effort to develop theory (Berente et al., 2019).

The theory developed by the authors indicates that activity variation in routine components leads to higher rates of successful merging of code (consequences). Activity variation and the other form of routine variation elicited from the data, order variation, were also shown to respond to different types of interdependencies (context). Activity variation was related to interdependencies across software code, while order variation was related to interdependencies across developers, indicating that routine components adjust to different circumstances. Indeed, the two types of routine components that were enacted seem to emerge from local adaptations to different circumstances (constitution). Hence, all three principles of pragmatism figured in the emergent theory and were also interrelated with each other. Therefore, an explanation was crafted that addressed the research problem from multiple, albeit integrated, perspectives. This allowed the authors to contribute to the literatures of online communities and to organization studies in general.

6.2 Tuertscher et al. (2014)

In this study the authors examined complex design, innovation and knowledge integration processes at the ATLAS/CERN physics research center. The problem confronting the authors was to explain how distributed actors can develop a novel, complex technological system despite the lack of hierarchy or a central coordinator. This problem is fundamentally concerned with interdependencies distributed across a vast system, and the authors thus collected a wide-ranging dataset that spanned 328 meetings, 84 semi-structured interviews, 128,015 items from electronic mailing lists, and 2,419 documents generated in the various meetings. All of these data pertained to the same overall coordination process, thus helping to achieve sample integration.

To address the research problem, drawing on the interview and archival data collected, the authors constructed a process model using qualitative analysis. This was complemented by using machine pattern recognition to first construct networks that indicated “interlaced knowledge,” and then show how patterns of “justification” correlated with higher degrees of interlaced knowledge, thus suggesting that justification is a key driver of coordination due to its boosting of interlaced knowledge. The human and machine pattern recognition thus corroborated each other through eliciting different aspects of the same overall coordination process.

The analyses cover multiple aspects of the phenomena at hand (i.e., forms of communication and distributions of knowledge) and the use of visualization connects these analyses. Figures 2-4 on pp. 16-17, for example, captures the relationships across different degrees of justification and interlaced knowledge. This shows how analyses conducted using both human and machine pattern recognition provide coverage across multiple aspects of the phenomena at hand as well as connections across such analyses.

In terms of balancing human and machine pattern recognition, this study leans mainly toward human pattern recognition performed using grounded theory. Machine pattern recognition (i.e., text mining and graph analysis) is mostly used in a confirmatory manner to corroborate the qualitative findings. This is a common pattern in mixed methods generally (Johnson, Onwuegbuzie, & Turner, 2007) as well as in contemporary IS research specifically (e.g., Leonardi, 2013).

The theory that emerges from this exercise is a process model (consequences) that centers around justification (i.e., arguments in favor of a particular position) and contestation (i.e., arguments challenging a particular position). Justification, however, shows different intensities in different workgroups, i.e., in situations where there are different levels of interlaced knowledge (context). Individual acts of justification combine to constitute a larger coordination process together with contestation (constitution). All the different aspects of the pragmatist principles thus add up to a theory with multiple interlocking parts. This theory addresses the research problem, which is concerned with coordination of interdependencies across a complex, heterogeneous system. The consideration of all three pragmatist principles therefore helps to craft a multifaceted explanation, thus providing the groundwork for the authors to make a contribution to the literatures on coordination of complex technological systems, knowledge creation and transformation, as well as innovation in distributed communities.

6.3 Vaast et al. (2017)

The authors of this study focused on the problem of explaining how social media affords connective action. To address this problem, this study used a dataset consisting of 23,000 tweets about the Gulf of Mexico oil spill. The choice of tweets as a data source, which contain text and are stamped with both time and actor information, laid the foundation for sample integration across multiple types of analysis pertinent to this research problem. The tweets were analyzed using both human and machine pattern recognition.

The paper has a complex structure and iterates back and forth between human and machine pattern recognition. The analysis focuses on "connective action episodes"—essentially connected sets of conversational actions. These episodes are narrated through grounded theory, and their evolution over time is visualized. Cluster analysis was used to identify groups of actors (i.e., microbloggers using Twitter) within the data. The topics of such conversational episodes were identified using grounded theory, and their changing intensity of participation over time was identified using visualizations. Finally, the authors triangulated collaboration patterns across actors through identifying patterns of interdependence across user groups. Through using motif analysis to identify different interaction patterns across actors, a relational structure (i.e., the social grouping or set of relationships in which an activity occurs) was used to enrich the qualitative observations made previously.

The analyses conducted thus provide coverage across multiple aspects of the phenomena (i.e., roles, episodes, and social structure motifs), while also establishing connections across these different analyses through the use of visualizations and cross-tabulations of social structure motifs across roles and episodes. For example, in Figures 5-7 on pp. 1194-1195, the authors show how different actors taking on different roles participate at different rates across time, for each of the different connective action episodes. This shows how different forms of analysis are connected through visualization.

In terms of the balance between human and machine pattern recognition, this study leans toward machine pattern recognition (i.e., graph analysis and visualization of temporal patterns). Human pattern recognition was performed using the grounded theory method to identify episodes that framed the overall analysis. This is a somewhat uncommon usage of qualitative analyses but speaks to the capacity of human pattern recognition to easily discern significance, i.e., the relationship of intended meaning to other concepts (Hirsch, 1967).

The theory developed by the authors focuses on showing how different affordances enable different types of connective action (consequences). This overall process was analyzed across multiple different episodes, thus showing how the process was active throughout multiple environments (context). Within each of these contexts, participants enacted different roles (constitution) that then participated in different interdependence relationships. Thus, all three pragmatist principles were integrated into the overall theorizing performed by the authors. This allowed the authors to address a multifaceted research problem and also contribute to multiple literatures: technology affordances, social media, and connective action.

7 Discussion

As the availability of digital trace data and computational tools increasingly enables modeling of a wide range of phenomena (Arazy et al., 2016; Burt, 2004; Johnson et al., 2015), opportunities for qualitative, theory-developing researchers to engage with such data and tools also increase. To this end, I have developed a pragmatist framework that indicates how human and machine pattern recognition can be used in an abductive fashion to generate new theory. From this framework, a number of tangible guidelines for how data, analysis, and theorizing can be handled emerge, as well as guidelines clarifying the evaluation of the process and product of such work.

Below, I first discuss the various ways in which machine pattern recognition can be appropriated by qualitative scholars. I then show how the approach suggested herein differs from other, related approaches and also reflect on the increasing convergence across human and machine pattern recognition. Finally, I discuss the persistent and unique role of human pattern recognition in theory development.

7.1 How Qualitative Researchers Can Appropriate Machine Pattern Recognition

The overall framework and guidelines provided in this paper elucidate general principles and practices that qualitative scholars can use to appropriate machine pattern recognition tools, techniques, and approaches. I focus here on a couple of ways in which such appropriation can occur, illustrated by empirical examples. These examples show how machine pattern recognition can be used by qualitative researchers to develop theory in three main ways: corroboration of structural patterns, exploratory data analysis, and construction of theoretical mechanisms.

The most obvious use, which is also traditionally employed in mixed methods studies (Johnson, Onwuegbuzie, & Turner, 2007), is the usage of machine pattern recognition to corroborate findings identified by qualitative methods. Often, this takes the form of identifying structural patterns that are implied by qualitative findings. For example, Tuertscher et al. (2014) identify a coordination process consisting of justification, contestation, and interlaced knowledge, and then use text mining and graph analysis to show how justification and interlaced knowledge correlate with each other. Such analysis can serve to provide additional corroboration of qualitative analyses. It can also serve as a way to show micro-macro linkages, thereby enabling zooming in and out (Gaskin et al., 2014). While qualitative findings reveal contextually embedded action, machine pattern recognition identifies structural patterns. Either way, the repertoire of qualitative researchers can be expanded through this approach to deliver findings corroborated by multiple methods, as well as analyses of structural patterns that may be difficult to achieve using traditional qualitative methods.

Machine pattern recognition may also be used as a means to explore large datasets in order to identify patterns that may be interesting for subsequent qualitative analysis (Zachariadis et al., 2013). For example, Lindberg et al. (2016) utilized visualizations to explore relationships among several variables germane to OSS development, and Vaast et al. (2017) used cluster analyses to elicit groupings and roles on Twitter. These analyses were then used to frame subsequent analysis. Qualitative analysts can therefore use machine pattern recognition to identify interesting variables and relationships that can provoke later qualitative investigations.

Finally, machine pattern recognition may be used as a means to construct theoretical mechanisms. This is not meant to imply that theory can be constructed "automatically," but rather that the increasing richness with which machine pattern recognition is able to discover and model patterns in text, processes, and relationships offers opportunities to interrelate multiple patterns and models arrived at by such means and use more complex interrelationships across variables and patterns as a basis for constructing theory (Berente et al., 2019.) This can, for example, be seen in Lindberg et al.'s (2016) usage of multiple visualizations and statistical models, which were then tied together to form the basis of an emergent theory. Similarly, Vaast et al. (2017) also drew together multiple computational analyses to paint a picture of both a set of changing relationships (role clusters and various relational motifs) and a set of processes (temporal distributions of role-related activity levels), thus providing the groundwork for the emergence of a theory.

These methods for qualitative researchers to appropriate machine pattern recognition may not be the only available means for doing so. They do, however, illustrate at least three important approaches through which qualitative researchers can extend their current capacities through interacting with digital trace data using computational tools in ways that contribute to their traditional goal of developing theory.

7.2 Comparing and Contrasting Other Approaches for Integrating Human and Machine Pattern Recognition

The pragmatist approach can be compared and contrasted to multiple other approaches already suggested in the IS literature. Note that the pragmatist approach is philosophically inclusive and does not engage directly with issues of ontology. Therefore, the approach suggested in this paper is complementary to these prior contributions and also inclusive of them. I discuss how the pragmatist framework developed herein differs in contribution from traditional mixed methods, critical realist studies, grounded theory, and sociomaterial studies using computational tools.

In the past, it was common to use mixed methods in a "sandwich-style" (Onwuegbuzie & Collins, 2007). Researchers would conduct a qualitative study and then use a traditional survey (with attendant structural equation modeling) to confirm or triangulate findings arrived at using human pattern recognition. Modern approaches to mixed methods (Brown et al., 2016; Venkatesh, Brown, & Bala, 2013) provide general frameworks for mixing traditional quantitative and qualitative methods in more sophisticated ways. These efforts are largely based on the premise of mixing disparate paradigms (e.g., mixing positivist approaches using quantitative methods with interpretive approaches using qualitative methods). In doing so, they often assume that quantitative methods are regression-based or econometric in nature, and that qualitative inquiry is mainly conducted on interview data. Because of this, they devote little specific attention to the idiosyncrasies of digital trace data and computational methods. In contrast, the iteration of human and machine pattern recognition across both discovery and justification functions under the umbrella of pragmatist principles, and therefore does not try to "mix" disparate paradigms. The pragmatist approach is therefore closer to the insistence of grounded theorists that "all is data" (Glaser, 2001) and also provides specific guidance on how to drive the abductive process forward. In summary, the pragmatist approach can help to achieve even tighter linkages between different forms of evidence, compared to what can be achieved using traditional mixed methods. Since the abductive approach integrates the different forms of data and analysis tightly in both the context of discovery and justification (Swedberg, 2012), a synthesis rather than a mix is achieved.

Similarly, multiple researchers (Wynn & Williams, 2012; Zachariadis et al., 2013) have utilized critical realism as an approach to mixed methods. These approaches have many similarities to the pragmatist approach, mostly in terms of their openness to multiple methods, and the retroductive approach to analysis, which, in many ways, is similar to the abductive approach espoused by pragmatist thinkers. Still, there are clear differences between their approach and what is suggested in this paper. First, critical realist approaches are centered around issues of structure and agency and use retroduction to uncover the unobservable, generative mechanisms that mediate between the agentic and structural levels. Second, while not specific to the critical realist approach in itself, neither Zachariadis et al. (2013) nor Wynn and Williams (2012) deal specifically with the nature of computational tools and their impact on the mixing of qualitative and quantitative methods. For example, Zachariadis et al. (2013, p. 862) specifically associate quantitative methods with econometrics. The approach spelled out here articulates how a variety of computational techniques can be used to analyze multiple, often structural, aspects of social action.

Qualitative researchers utilizing grounded theory have, in the last few years, begun to grapple with some of the issues discussed in this paper (Walsh et al., 2015b). Most of these attempts have been efforts to move beyond interview transcripts as the main source of data, and make more effective use of digital traces of various kinds (Birks, Fernandez, Levina, & Nasirin, 2013; Gasson & Waters, 2011; Vaast & Walsham, 2011; Walsh et al., 2015a). This engagement, however, has mostly focused on digital traces rather than on computational tools, and on methodology as opposed to providing a philosophical basis for such work. Some recent contributions, most notably Berente et al. (2019), have focused explicitly on showing how computational tools can be integrated into the grounded theory paradigm and how researchers must continuously make sense of various forms of data and attendant analyses. The pragmatist framework contributes to this line of work through providing philosophical backing, which also allows for eliciting tangible guiding principles for how such sensemaking can be conducted. Through drawing on the three pragmatist principles elucidated in this paper, scholars are provided with specific guidance for how to approach the abductive process of discovering and justifying working hypotheses, using both human and machine pattern recognition methods.

Finally, Gaskin et al. (2014) propose a computational approach for analyzing sociomaterial routines using a combination of content analysis and sequence analysis. This approach is highly tailored to a specific context (i.e., design using digital tools) and a particular level of analysis (i.e., routines). They argue that their framework contrasts with traditional, qualitative methods for studying practices, which tend to emphasize the local, contextual, and idiosyncratic actions of individual agents (Gaskin et al., 2014, p. 863). Hence, Gaskin et al.'s proposed approach tends to be more effective for uncovering structural patterns, as opposed to analyzing agentic dispositions. The pragmatist approach, in contrast, provides broad guidelines for how to integrate human and machine pattern recognition, enabling the analysis of a wide range of phenomena beyond design or routines, while paying attention to constitution, context, and consequences. Rather than challenge the assumptions and assertions made by Gaskin et al. (2014), I argue that their approach represents a special case of the pragmatist method presented in this paper.

7.3 The Convergence of Data and Methods for Human and Machine Pattern Recognition

The notion that qualitative and quantitative inquiry represent different ontologies or epistemologies is becoming increasingly tenuous. This division largely seems to stem from what some have labeled "extensive" and "intensive" properties of phenomena (DeLanda, 2005). Extensive properties are those that have an extension in space or time, such as physical size or duration. Other properties, however, are "intensive," meaning that they do not refer to the extent of something, but rather to their degree of intensity. In the social realm, many properties of interest have no extension, but differ in their degree of intensity. For example, emotions can vary in their intensity and so can the influence of various social structures. Intensities, however, can also increasingly be measured.

Recognizing both extensive and intensive properties, the pragmatist perspective works in concert with new forms of data and analytical methods to help further narrow an already rapidly closing gap between human and machine pattern recognition methods. Pragmatism helps contribute to this process through its focus on action, which fits well with increasingly available digital trace data that can be conveniently analyzed both in qualitative and quantitative terms. For example, digital trace data in the form of text is increasingly used in social science research, and there is a need to understand how to work with such data to a greater degree, as it is quite possible that such data will increasingly compete with interviews as our main source of qualitative data (DiMaggio, 2015; Grimmer & Stewart, 2013). Further, using the pragmatist framework, researchers can utilize methods that allow for the quantification of processes and narratives in ways that are less reductive and more sensitive to various contextual factors, compared to prior generations of quantitative tools. In summary, a richer view of how to approach intensive properties is suggested by the pragmatist approach elucidated in this paper.

Integrating analyses of extensive and intensive properties using the pragmatist approach offers opportunities to develop more novel and interesting theories. Previously, human pattern recognition (i.e., traditional qualitative methods) has often been used to craft "explanatory" theory, whereas machine pattern recognition, especially as manifested by machine learning and other "black-boxed" approaches, has often been used to create "predictive" theory (Gregor, 2006). The pragmatist framework presented in this paper breaks down such old stereotypes by showing how human and machine pattern recognition can be used to develop theory that is both explanatory and predictive, without necessarily assigning the role of generating explanations through the examination of intensive properties to human pattern recognition or the role of generating predictions through the examination of extensive properties to machine pattern recognition (Lee, 1991). This provides opportunities to develop theories that have richer support in multiple forms of data and analysis, thus allowing for the integration of extensive and intensive properties under the shared umbrella of the pragmatist approach. Such theories may be novel and interesting (Davis, 1971) because data and analyses can be used in surprising ways—for example, through using human pattern recognition to support predictions or machine pattern recognition to support explanations.

7.4 The Unique Role of Human Pattern Recognition in Theory Development

While parts of the pattern recognition aspects of theoretical development work can be automated (Berente et al., 2019), suggesting a supplementary role for computational tools, the process of interrelating variegated patterns generated by disparate methods to develop theory needs the essential component of human, disciplined imagination (Cornelissen, 2006; Weick, 1989) to identify additional working hypotheses that contain possible correlations, patterns, and explanations for why, in particular, other patterns occur.

For example, developing software ecosystems involves technical, social, political, and economic issues that must be negotiated, interleaved, and resolved across long stretches of time, thereby necessitating the application of heterogeneous knowledge resources and coordination of disparate groups and organizations (Lehman, 1980). Understanding and intervening in such complex systems is likely to require not only pattern recognition capacities, but also the imaginative capacities of the human mind. The heart of the scientific enterprise is thus situated in human pattern recognition and its capacity to create coherent accounts of multiple patterns, whether they are identified by humans or machines, or explained by extant or emergent theory.

Human pattern recognition, as opposed to machine pattern recognition, is closely linked to creativity, intuition, and the ability to forge wide-ranging connections between disparate forms of data and analysis. This represents a fulcrum around which the role of humans in the theory development process revolves. In some respects, i.e., the raw identification of patterns in delimited datasets, the gap between machine and human pattern recognition diminishes as computational tools grow more sophisticated. In contrast, the capacity to tie together disparate empirical patterns into coherent theoretical accounts that can be contrasted with past findings and theories is a unique, innate capacity of the human mind. All of these aspects are central to theory development as it is conducted today in management and information systems research and are difficult for machines to replicate. Therefore, I argue that human pattern recognition has a deep and unique role in the theory development activities that are central to our progress as a discipline and our capacity to generate results that may hold important lessons for practitioners.

Qualitative researchers, therefore, can leverage human pattern recognition in combination with computational tools to conduct analyses of large-scale patterns in massive digital trace datasets. This does not amount to a conversion to the ideology of “confirmatoids” (Dougherty, 2015), but rather reflects an increasing engagement between human and machine pattern recognition:

> *the new climate in AI favors systems that advise humans rather than replace them, and recent analyses of machine learning applications (e.g., Langley & Simon, 1995) suggest an important role for the developer. Such analysis carry over directly to discovery in scientific domains* (Langley, 2000, p. 396)

Thus, while human pattern recognition finds itself at the core of the theory development process, the toolbox used to identify the patterns that form the basis of theory can evolve to encompass new advances in data and analytical tools, without compromising on the intent to generate strong theoretical accounts of human affairs. This helps disentangle the various parts of the theory development process in order to identify the ways in which it can and cannot be augmented by computational tools (Grimmer & Stewart, 2013).

7.5 Conclusion

Digital traces and computational tools offer important opportunities to qualitative researchers engaged in theory development. These tools are increasingly both dynamic and inductive and are therefore consistent with the traditional proclivities and interests of qualitative researchers. In closing, I argue that it is possible for researchers to engage in research projects that retain the traditional, qualitative goal of developing theory, while also embracing the new opportunities that digital trace data and computational tools offer.

Acknowledgments

I would like to thank the following people who provided helpful comments on earlier drafts: James Gaskin, Nicholas Berente, Hani Safadi, Yan Chen, and Kalle Lyytinen. I have also benefited from seminar discussions at the University of Georgia, New York University, and the Doctoral Seminar on Research Quality in Information Systems in Davos, Switzerland.

References

Abbott, A. (1992). From causes to events: Notes on narrative positivism. *Sociological Methods & Research*, *20*(4), 428-455.

Agar, M. (1985). *Speaking of ethnography*. SAGE.

Aldrich, H., & Ruef, M. (2006). *Organizations evolving*. SAGE.

Alvesson, M., & Sandberg, J. (2011). Generating research questions through problematization. *Academy of Management Review, 36*(2), 247-271.

Arazy, O., Daxenberger, J., Lifshitz-assaf, H., Nov, O., & Gurevych, I. (2016). Turbulent stability of emergent roles: The dualistic nature of self-organizing knowledge co-production. *Information Systems Research, 24*(7), 792-812.

Arazy, O., Nov, O., Patterson, R., & Yeo, L. (2011). Information quality in Wikipedia: The effects of group composition and task conflict. *Journal of Management Information Systems, 27*(4), 71-98.

Ayim, M. (1974). Retroduction: The rational instinct. *Transactions of the Charles S. Peirce Society, 10*(1), 34-43.

Baldwin, J. D. (1988). Habit, emotion, and self-conscious action. *Sociological Perspectives, 31*(1), 35-57.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Behfar, K., & Okhuysen, G. A. (2018). Perspective—Discovery within validation logic: Deliberately surfacing, complementing, and substituting abductive reasoning in hypothetico-deductive inquiry. *Organization Science*, orsc.2017.1193.

Berente, N., Seidel, S., & Safadi, H. (2019). Research commentary: Data-driven computationally-intensive theory development. *Information Systems Research, 30*(1), 50-64.

Berger, J., & Pope, D. (2011). Can losing lead to winning? *Management Science, 57*(5), 817-827.

Berger, P. L., & Luckmann, T. (1967). *The social construction of reality: A treatise in the sociology of knowledge*. Anchor.

Birks, D. F., Fernandez, W., Levina, N., & Nasirin, S. (2013). Grounded theory method in information systems research: Its nature, diversity and opportunities. *European Journal of Information Systems, 22*(1), 1-8.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America, 99* Suppl 3, 7280-7287.

Brown, S. S. A., Venkatesh, V., Brown, S. S. A., & Sullivan, Y. (2016). Guidelines for conducting mixed-methods research: An extension and illustration. *Journal of the Association for Information Systems, 17*(7), 435-495.

Bühlmann, P., Wyner, A. J., Buhlmann, P., & Wyner, A. J. (1999). Variable Length Markov Chains. *The Annals of Statistics, 27*(2), 480-513.

Burks, A. (1946). Peirce's theory of signs. Philosophy of Science, 13(4), 301-306.

Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology, 110*(2), 349-399.

Carlile, P. R. P. (2002). A pragmatic view of knowledge and boundaries: Boundary objects in new product development. *Organization Science*, *13*(4), 442-455.

Charmaz, K. (2014). *Constructing grounded theory*. SAGE Publications.

Chiasson, P. (2007). The Role of optimism in abduction. *Commens Digital Companion to C. S. Peirce*, http://www.commens.org/encyclopedia/article/chiasson-phyllis-role-optimism-abduction

Cohen, M. D. (2007). Reading Dewey: Reflections on the study of routine. *Organization Studies, 28*(5), 773-786.

Cohen, M.D., Burkhart, R., Dosi, G., Egidi, M., Marengo, L., Warglien, M., & Winter, S. (1996). Routines and other recurring action patterns of organizations: Contemporary research issues. *Industrial and Corporate Change, 5*(3), 653.

Cohen, Michael D. (2012). Perceiving and remembering routine action: Fundamental micro-level origins. *Journal of Management Studies, 49*(8), 1383-1388.

Contractor, N., & Monge, P. (2011). Multidimensional networks and the dynamics of sociomateriality: Bringing technology inside the network. *International Journal of Communication, 5*, 682-720.

Cornelissen, J. P. (2006). Making sense of theory construction: Metaphor and disciplined imagination. *Organization Studies, 27*(11), 1579.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302.

Davis, M. S. (1971). That's Interesting! Towards a phenomenology of sociology and a sociology of phenomenology. *Philosophy of the Social Sciences*, *1*(2), 309-344.

DeLanda, M. (2005). *Intensive science & virtual philosophy*. Bloomsbury Academic.

Dewey, J. (1910). *How we think*. Heath & Co.

Dewey, J. (1938a). *Logic: The theory of inquiry*. Henry Holt.

Dewey, J. (1938b). The pattern of inquiry. In *Logic: The theory of inquiry*. Henry Holt.

Dhar, V., & Chou, D. (2001). A comparison of nonlinear methods for predicting earnings surprises and returns. *IEEE Transactions on Neural Networks, 12*(4), 907-921.

Dhar, Vasant. (2013). Data science and prediction. *Communications of the ACM, 56*(12), 64-73.

DiMaggio, P. (2015). Adapting computational text analysis to social science (and vice versa). *Big Data & Society, 2*(2), 205395171560290.

Dougherty, D. (2015). Reflecting on the reflective conversation. *Organizational Research Methods, 18*(4), 606-611.

Dougherty, D. (2016). *Taking advantage of emergence: Productively innovating in complex innovation systems.* Oxford University Press.

Džeroski, S., Langley, P., & Todorovski, L. (2007). *Computational discovery of scientific knowledge*. Springer

Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic capabilities: What are they? *Strategic Management Journal, 21*(10-11), 1105-1121.

Elder-Vass, D. (2011). The *causal power of social structures: Emergence, structure and agency.* Cambridge University Press.

Faraj, S., von Krogh, G., Monteiro, E., & Lakhani, K. R. (2016). Online community as space for knowledge flows. *Information Systems Research, 27*(4), 668-684.

Farjoun, M., Ansell, C., & Boin, A. (2015). Perspective—Pragmatism in organization studies: Meeting the challenges of a dynamic and complex world. *Organization Science, 26*(6), 1787-1804.

Fayyad, U. M., Grinstein, G. G., & Wierse, A. (2002). *Information visualization in data mining and knowledge discovery*. MK/Morgan Kaufmann Publishers.

Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly, 48*(1), 94-121.

Gaskin, J., Berente, N., Lyytinen, K., & Yoo, Y. (2014). Toward generalizable sociomaterial inquiry: A computational approach for zooming in and out of sociomaterial routines. *MIS Quarterly, 38*(3), 849-871.

Gasson, S., & Waters, J. (2011). Using a grounded theory approach to study online collaboration behaviors. *European Journal of Information Systems, 22*(1), 95-118.

Geertz, C. (1973). *The Interpretation of cultures: Selected essays*. Basic Books.

Giddens, A. (1984). The Constitution of Society: Outline of the theory of structuration. University of California Press.

Glaser, B. G. (2001). *The grounded theory perspective: Conceptualization contrasted with description*. Sociology Press.

Goertz, G. (2006). *Social science concepts: A user's guide*. Princeton University Press.

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology, 91*(3), 481-510.

Gregor, S. (2006). The Nature of theory in information systems. *MIS Quarterly, 30*(3), 611-642.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis, 21*(3), 267-297.

Grolemund, G., & Wickham, H. (2014). A Cognitive interpretation of data analysis. *International Journal of Statistics, 82*(2), 184-204.

Gronow, A. (2012). From habits to social institutions: A pragmatist perspective. *Studies across Disciplines in the Humanities and Social Sciences*, *12*, 26-44.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis*. Pearson Education.

Hedman, J., Srinivasan, N., & Lindgren, R. (2013). Digital Traces of information systems: Sociomateriality made researchable. *Proceedings of the 34th International Conference on Information Systems*.

Hirsch, E. D. (1967). *Validity in Interpretation.* Yale University Press.

Holland, J. H. (1992). Complex adaptive systems. *Daedalus, 121*(1), 17-30.

Holland, J., Holyoak, K., Nisbett, R., & Thagard, P. (1989). Induction: Processes of inference, learning, and discovery. MIT Press.

Howison, J., & Crowston, K. (2014). Collaboration through open superposition: A theory of the open source way. *MIS Quarterly, 38*(1), 29-50.

Howison, J., Wiggins, A., & Crowston, K. (2011). Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems, 12*(12), 767-797.

James, W. (1907). Pragmatism: A new name for some old ways of thinking. Lectures on philosophy. Longmans, Green & Co.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research, 1*(2), 112-133.

Johnson, S. L., Safadi, H., & Faraj, S. (2015). The emergence of online community leadership. *Information Systems Research, 26*(1), 165-187.

Kaufman, L., & Rousseeuw, P. J. (2005). Finding groups in data: An introduction to cluster analysis. John Wiley & Sons.

Ketoviki, M., & Mantere, S. (2010). Two strategies for inductive reasoning in organizational research. *Academy of Management Review, 35*(2), 315-333.

Langley, A. (1999). Strategies for theorizing from process data. *Academy of Management Review, 24*(4), 691-710.

Langley, P. (2000). Computational support of scientific discovery. *International Journal of Human Computer Studies, 53*(3), 393-410.

Latour, B. (2006). *Reassembling the social.* Oxford University Press.

Lee, A. (1991). Integrating positivist and interpretive approaches to organizational research. *Organization Science, 2*(4), 342-365.

Lee, A. A. S. A., & Baskerville, R. R. L. R. (2003). Generalizing generalizability in information systems research. *Information Systems Research, 14*(3), 221-243.

Lehman, M. (1980). Programs, life cycles, and laws of software evolution. *Proceedings of the IEEE*, 68(9), 1060-1076.

Leonardi, P. (2013). When does technology use enable network change in organizations? A comparative study of feature use and shared affordances. *MIS Quarterly, 37*(3), 749-775.

Lindberg, A, Gaskin, J., Berente, N., Lyytinen, K., & Yoo, Y. (2013). Computational approaches for analyzing latent social structures in open source organizing. *Proceedings of the 34th International Conference on Information Systems*.

Lindberg, Aron, Berente, N. N., Gaskin, J., & Lyytinen, K. (2016). Coordinating Interdependencies in online communities: A study of an open source software project. *Information Systems Research, 27*(4), 751-772.

Locke, K., Golden-Biddle, K., & Feldman, M. S. (2008). Making doubt generative: Rethinking the role of doubt in the research process. *Organization Science*, *19*(6), 907-918.

Locke, Karen, & Golden-Biddle, K. (1997). Constructing opportunities for contribution: structuring intertextual coherence and “problematizing” in organizational studies. *Academy of Management Journal, 40*(5), 1023-1062.

Miranda, S. M., Kim, I., & Summers, J. D. (2015). Jamming with social media: Jow cognitive structuring of organizing vision facets affects IT innovation diffusion. *MIS Quarterly, 39*(3), 591-614.

Moretti, F. (2013). *Distant reading.* Verso.

Onwuegbuzie, A., & Collins, K. (2007). A typology of mixed methods sampling designs in social science research. *The Qualitative Report, 12*(2), 281-316.

Onwuegbuzie, A. J., & Johnson, R. B. (2006). The validity issue in mixed research. *Research in the Schools*, *13*(1), 48-63.

Orlikowski, W. J. W. J. (2000). Using technology and constituting structures: A practice lens for studying technology in organizations. *Organization Science, 11*(4), 404-428.

Ormerod, R. (2006). The history and ideas of pragmatism. *Journal of the Operational Research Society, 57*(8), 892-909.

Paavola, S. (2005). Peircean abduction: Instinct or inference? *Semiotica, 153*, 131-154.

Pentland, B. T. (2003). Conceptualizing and measuring variety in the execution of organizational work processes. *Management Science, 49*(7), 857-870.

Pollock, N., & Williams, R. (2008). *Software and organisations: The biography of the enterprise-*

wide system or how SAP conquered the world. Taylor & Francis.

Popper, K. (1965). Three views concerning human knowledge. In *Conjectures and refutations: The growth of scientific knowledge* (pp. 97-119). Routledge

Ragin, C. C., & Strand, S. I. (2008). Using qualitative comparative analysis to study causal order: comment on Caren and Panofsky (2005). *Sociological Methods & Research, 36*(4), 431-441.

Rai, A. (2017). Avoiding type III errors: Formulating IS research problems that matter. *MIS Quarterly, 41*(2), iii-vii.

Rescher, N. (2000). *Realistic pragmatism: An introduction to pragmatic philosophy*. State University of New York Press.

Sami, P. (2004). Abduction as a logic and methodology of discovery: The importance of strategies. *Foundations of Science, 9*(3), 267-283.

Savigny, E. von, Schatzki, T., & von Savigny, E. (2001). *The practice turn in contemporary theory*. Routledge.

Searle, J. R. (1969). *Speech acts: an essay in the philosophy of language*. Cambridge University Press.

Sewell, Jr., W. H. (1992). A theory of structure: Duality, agency, and transformation. *American Journal of Sociology, 98*(1), 1-29.

Shmueli, G., & Koppius, R. O. (2011). Predictive analytics in information systems research. *MIS Quarterly*, *35*(3), 553-572.

Simpson, B. (2009). Pragmatism, Mead, and the practice turn. *Organization Studies*, *30*(12), 1329-1347.

Strauss, A., & Corbin, J. M. (1990). *Basics of qualitative research: grounded theory procedures and techniques.* SAGE.

Swedberg, R. (2012). Theorizing in sociology and social science: Turning to the context of discovery. *Theory and Society, 41*(1), 1-40.

Tubaro, P., & Casilli, A. (2010). “An ethnographic seduction”: How qualitative research and agent-based models can benefit each other. *Bulletin de Méthodologie Sociologique, 106*(1), 59-74.

Tuertscher, P., Garud, R., & Kumaraswamy, A. (2014). Justification and interlaced knowledge at ATLAS, CERN. *Organization Science*, *25*(6), 1579-1608.

Tufte, E. R. (2001). *The visual display of quantitative information*. Graphics Press.

Tukey, J. W. (1977). *Exploratory data analysis.* Addison-Wesley.

Vaast, E., Safadi, H., Lapointe, L., & Negoita, B. (2017). Social media affordances for connective action: An examination of microblogging use during the Gulf of Mexico oil spill. *MIS Quarterly, 41*(4), 1179-1205.

Vaast, E., & Walsham, G. (2011). Grounded theorizing for electronically mediated social contexts. *European Journal of Information Systems, 22*(1), 9-25.

van der Aalst, W., Adriansyah, A., De Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., … Wynn, M. (2011). Process Mining Manifesto. *Proceedings of the BPM 2011 International Workshops*.

Venkatesh, V., Brown, S., & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly, 37*(1), 21-54.

Venturini, T., & Latour, B. (2010). The social fabric: Digital traces and quali-quantitative methods. *Proceedings of Future En Seine*.

Walsh, I. (2015). Using quantitative data in mixed-design grounded theory studies: An enhanced path to formal grounded theory in information systems. *European Journal of Information Systems*, *24*(5), 531-557.

Walsh, I., Holton, J. A., Bailyn, L., Fernandez, W., Levina, N., & Glaser, B. (2015a). Rejoinder: moving the management field forward. *Organizational Research Methods*, *18*(4), 620–628.

Walsh, I., Holton, J. A, Bailyn, L., Fernandez, W., Levina, N., & Glaser, B. (2015b). What grounded theory is: A critically reflective conversation among scholars. *Organizational Research Methods, 18*(4), 581-599.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications.* Cambridge University Press.

Webb, E., & Weick, K. (1979). Unobtrusive measures in organizational theory: A reminder. *Administrative Science Quarterly*, *24*(4), 650-659.

Weick, K. E. (1979). *The social psychology of organizing* (Vol. 2). Addison-Wesley.

Weick, K. E. (1989). Theory construction as disciplined imagination. Academy of Management Review, 14(4), 516-531.

Weick, K. E. (1995). Editing innovation into *Administrative Science Quarterly*. In L. L. Cummings & P. J. Frost (Eds.), *Publishing in the organizational sciences*. SAGE.

Whelan, E., Teigland, R., Vaast, E., & Butler, B. (2016). Expanding the horizons of digital social networks: Mixing big trace datasets with qualitative approaches. *Information and Organization*, *26*(1-2), 1-12.

Whitford, J., & Zirpoli, F. (2014). Pragmatism, practice, and the boundaries of organization. *Organization Science, 25*(6), 1823-1839.

Winter, S. G. (2013). Habit, deliberation, and action: Strengthening the microfoundations of routines and capabilities. *Academy of Management Perspectives*, *27*(2), 120-137.

Wynn, D., & Williams, C. K. (2012). Principles for conducting critical realist case study research in information systems. *MIS Quarterly, 36*(3), 787-810.

Zachariadis, M., Scott, S., & Barrett, M. (2013). Methodological implications of critical realism for mixed-methods research. *MIS Quarterly, 37*(3), 855-879.

About the Author

Aron Lindberg is an assistant professor of information systems in the School of Business at Stevens Institute of Technology. He received his PhD at Case Western Reserve University (2015), and primarily studies complex design processes, often using a combination of qualitative and computational methods. His research has been published or is forthcoming in journals such as *MIS Quarterly*, *Information Systems Research*, and *Journal of the Association for Information Systems.*